Prognosis and Gene Expression Profiling of 20q13-Amplified Breast Cancers

Christophe Ginestier,[1] Nathalie Cervera,[1] Pascal Finetti,[1] Séverine Esteyries,[1] Benjamin Esterni,[2] José Adélaïde,[1] Luc Xerri,[3] Patrice Viens,[2,4] Jocelyne Jacquemier,[1,3] Emmanuelle Charafe-Jauffret,[1,3,4] Max Chaffanet,[1] Daniel Birnbaum,[1] and François Bertucci[1,2,4]

Abstract

Purpose: Amplification of chromosomal region 20q13 occurs in breast cancer but remains poorly characterized.

Experimental Design: To establish the frequency of 20q13 amplification and select the amplified cases to be studied, we used fluorescence *in situ* hybridization of bacterial artificial chromosome probes for three 20q13 loci (*MYBL2, STK6, ZNF217*) on sections of tissue microarrays containing 466 primary carcinoma samples. We used Affymetryx whole-genome DNA microarrays to establish the gene expression profiles of 20q13-amplified tumors and quantitative reverse transcription-PCR to validate the results.

Results: We found 36 (8%) 20q13-amplified samples. They were distributed in two types: type 1 tumors showed *ZNF217* amplification only, whereas type 2 tumors showed amplification at two or three loci. Examination of the histoclinical features of the amplified tumors showed two strikingly opposite data. First, type 1 tumors were more frequently lymph node – negative tumors but were paradoxically associated with a poor prognosis. Second, type 2 tumors were more frequently lymph node – positive tumors but were paradoxically associated with a good prognosis. Type 1 and type 2 showed different gene expression profiles. No 20q13 gene could be associated with type 1 amplification, whereas several 20q13 genes were overexpressed in type 2 tumors.

Conclusions: Our results suggest that amplified tumors of types 1 and 2 are two distinct entities resulting from two different mechanisms and associated to different prognosis.

Genome alterations and aneuploidy are frequent in breast cancers. Several chromosomal regions of amplification have been identified and characterized. The most frequent are 8p12, 8q24, 11q13, and 17q12 (1). Each is found in 10% to 25% of cases. Amplification at 20q13 is slightly less frequent and occurs in around 5% to 12% of cases. The biological processes that lead to DNA amplification are poorly understood. A theory postulates that amplification of a gene with oncogenic activity when overexpressed is positively selected and is accompanied by that of neighbor genes as neutral passengers. In this case, amplification of the oncogene locus is supposed to correlate with overexpression of the corresponding mRNA and protein. This seems to be the case for the 17q12 alteration; it is generally admitted that the *ERBB2* gene, which encodes a tyrosine kinase receptor, drives the 17q12 amplification unit (amplicon). Using DNA microarrays, we have recently described a gene expression signature (GES) of the 17q12 amplification that is in agreement with this theory (2). For the other amplifications, it is more complicated to distinguish the oncogenes involved in the absence of functional validation of candidates and because the amplification unit is often made of several distinct amplicons. It is not known for certain whether there are one or several selected genes. It is probable that the *CCND1* gene, which encodes G_1-S cyclin, is a driver gene of at least some of the 11q13 amplicons. Several genes seem to be involved in the 8p12 amplification (3–5).

There are three major interests to study DNA amplification. First, this type of alteration is often associated with a specific prognosis or biological factor. Thus, amplification of *ERBB2* is a factor of poor prognosis, and that of *CCND1* is more frequent in tumors with positive estrogen receptor (ER) status. Second, we may learn on the mechanisms of oncogenesis from the function of genes involved. Third, identification of an oncogene may allow targeted therapy, such as the use of trastuzumab to treat *ERBB2*-amplified tumors.

Our knowledge of the 20q13 region amplification is limited. Several subregions and potential drivers have been suggested (6), including *STK6/AURKA*, which encodes the Aurora A serine/

Authors' Affiliations: [1]Laboratoire d'Oncologie Moléculaire, Centre de Recherche en Cancérologie de Marseille, UMR599 Inserm; [2]Département d'Oncologie Médicale et Investigation Clinique; [3]Laboratoire de BioPathologie, Institut Paoli-Calmettes; and [4]Université de la Méditerranée, UFR de Médecine, Marseilles, France
Received 10/27/05; revised 2/27/06; accepted 4/19/06.

Grant support: Inserm, Institut Paoli-Calmettes, Ligue Nationale Contre le Cancer (label 2003-2006) and Ministries of Health and Research (Cancéropôle PACA), and a fellowship from Ministry of Research (C. Ginestier and S. Esteyries).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

C. Ginestier and N. Cervera contributed equally to this work.

Requests for reprints: Daniel Birnbaum, UMR599 Inserm, 27 Bd. Leï Roure, 13009 Marseilles, France. Phone: 33-4-91-75-84-07; Fax: 33-4-91-26-03-64; E-mail: birnbaum@marseille.inserm.fr.

doi:10.1158/1078-0432.CCR-05-2339

threonine kinase, and *ZNF217* and *MYBL2*, which both code for transcription factors. However, despite thorough analyses (7), no definite conclusion has been drawn. Aneuploidy and aggressive behavior are often associated with 20q13 amplification (8, 9). Aneuploidy is likely to be due to overexpression of Aurora A, which is known to regulate mitosis (10, 11).

We studied three potential driver genes of the 20q13 amplification using fluorescence *in situ* hybridization (FISH) of locus-specific bacterial artificial chromosome (BAC) probes on sections of 466 breast tumors arrayed in tissue microarrays (TMA). We studied the relation between 20q13 amplification and histoclinical factors. We then used whole-genome DNA microarrays to define the gene expression profile of tumors with 20q13 amplification.

Patients, Materials, and Methods

Patients and samples. A consecutive series of samples from 685 women with unilateral localized invasive breast carcinomas treated at the Institut Paoli-Calmettes between June 1981 and December 2000 were included in TMAs. Of the 685 cases, 466 were available for both FISH and immunohistochemistry. For the gene expression analysis, we selected 55 samples on the basis of availability of good-quality RNA, including 19 of the 36 samples identified as amplified by FISH on TMA and 36 identified as nonamplified by FISH on TMA. The main histoclinical features of the tumors are summarized in Supplementary Table S1.

Tissue microarrays. TMAs were prepared as previously described (12). For each case, three representative areas from the tumor were carefully selected from a H&E–safran–stained section of a donor block. Core cylinders (0.6-mm diameter) were punched from each of them and deposited into three separate recipient paraffin blocks using a specific arraying device (Beecher Instruments, Silver Spring, MD). In addition to tumor samples, the recipient block also received internal controls including 10 normal breast tissue samples from 10 healthy women that underwent reductive mammary surgery, and pellets from various cell lines. Five-micrometer sections of the resulting TMA blocks were made and used for FISH and immunohistochemical analysis after transfer onto glass slides.

FISH on TMA analysis. To characterize the 20q13 amplification, dual-color FISH analysis was directly done on tumor samples arrayed in TMA, according to published protocols (13, 14). Three 20q13 subregions of amplification corresponding to *MYBL2*, *STK6*, and *ZNF217* loci were analyzed with locus-specific BAC pools. Based on the split-signal FISH approach (15), we used a combination of two differentially labeled pools of BAC: biotinylated locus-specific BAC pools (revealed in green, FITC) and digoxigenin-labeled BAC pools (revealed in red, TRITC) for centromere 20 region as a reference. From telomere to centromere, the different BAC pools were constituted as follows: BAC pool 1 (*STK6* locus): RP11-380D15 (AL139824; chr20:55,374,926-55,568,069), RP5-1167H4 (AL121914; chr20: 55,588,473-55,724,165), RP5-1153D9 (AL109806; chr20: 55,724,066-55,818,186); BAC pool 2 (*ZNF217* locus): RP11-91L1(chr20: 52,673,232-52,824,844), RP4-724E16 (AL157838; chr20: 52,813,526-52,942,378), RP11-299C12 (chr20: 52,899,287-53,089,979); BAC pool 3 (*MYBL2* locus): RP11-69I10 (chr20: 42,719,098-42,883,748), RP11-153L9 (chr20: 42,911,471-43,060,531), RP5-1030M6 (AL035089; chr20: 43,068,183-43,241,986); BAC pool 4 (centromere 20), RP11-243J16 (AL160175; chr20: 31,038,018-31,206,877), RP1-310O13 (AL031658; chr20: 31,220,458-31,383,433), RP5-836N17 (AL049539; chr20: 31,408,244-31,519,937), RP5-836N17 (AL121897; chr20: 31,519,838-31,665,243). Genomic information was taken from the University of California San Francisco Genome Browser on Human (May 2004 Assembly), which is based on National Center for Biotechnology Information Build 35 (National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda, MD). DNA from BAC clones was purified, labeled, and individually verified for its specificity for chromosome 20. All BAC clones were obtained from the BACPAC resource (Children's Hospital Oakland-BACPAC Resources, Oakland, CA). After counterstaining with Vectashield containing 4,6-diamidino-2-phenylindole (Vector Laboratories, Burlingame, CA), images were analyzed with a microscope (DMRXA, Leica Microsystèmes, Marseilles, France), captured with a charge coupled device camera, filtered, and processed with ISIS software (*In situ* Imaging Systems, Metasystems Hard-und Software GmbH, Altlussheim, Germany).[5] Areas enriched in tumor cells were identified by reference to near-adjacent sections stained with H&E–safran, and fluorescence was scored on a minimum of 50 nuclei per tumor. Two observers (CG, NC) read the TMAs independently. A ratio of locus-specific BAC pool signal/centromere 20 BAC pool signal was calculated for each locus analyzed. The locus analyzed was considered as amplified when this ratio was >4.0 in the cell. Tumors were defined as amplified when 10% or more of tumor cells showed such amplification.

Immunohistochemical analysis. Protein expression status of 685 tumor samples was done by immunohistochemistry on TMA for ER, PR, ERBB2, Ki67 and STK6/Aurora A. The characteristics of the antibodies used are listed in Table 1. All were commercial antibodies except anti-Aurora A, which was a gift from C. Prigent (16). Controls for immunohistochemical analysis have been previously described (17). Immunohistochemistry was carried out on 5 μm sections of tissue fixed in alcohol formalin for 24 hours and included in paraffin. Sections were deparafinized in Histolemon (Carlo Erba Reagenti, Rodano, Italy) and rehydrated in graded alcohol. Antigen enhancement was done by incubating the sections in target retrieval solution (Dako, Copenhagen, Denmark) as recommended. The reactions were carried out using an automatic stainer (Dako Autostainer). Staining was done at room temperature as follows: after washes in phosphate buffer, followed by quenching of endogenous peroxidase activity by treatment with 0.1% H_2O_2, slides were first incubated with blocking serum (Dako) for 30 minutes and then with the affinity-purified antibody for 1 hour. After washes, slides were incubated with biotinylated antibody against rabbit immunoglobulin for 20 minutes followed by streptavidin-conjugated peroxidase (Dako LSABR2 kit). Diaminobenzidine or 3-amino-9-ethylcarbazole was used as the chromogen. Slides were counterstained with hematoxylin, and coverslipped using Aquatex (Merck, Darmstadt, Germany) mounting solution. They were evaluated under a light microscope by two pathologists (E.C.J. and J.J.). The results were expressed in terms of percentage (P) and intensity (I) of positive cells as described previously (12, 17) and results were scored by the quick score (Q; $Q = P \times I$). For the TMA, the mean of the score of two core biopsies minimum was done for each case.

RNA extraction. Total RNA was extracted from frozen samples as previously described (18) and integrity was controlled by denaturing formaldehyde agarose gel electrophoresis and by microanalysis (Agilent Bioanalyzer, Palo Alto, CA).

DNA microarrays hybridizations. Nineteen of the 36 primary breast tumors identified as amplified by FISH on TMA and 36 identified as nonamplified by FISH on TMA were analyzed for gene expression using Affymetrix U133 Plus 2.0 human oligonucleotide microarrays, as recommended by the supplier.[6] Briefly, for each sample, synthesis of the first-strand cDNA was done from 3 μg total RNA by T7-oligo(dT) priming, followed by second-strand cDNA synthesis. After purification of cDNA, an *in vitro* transcription combined with amplification of cRNA was used to generate the cRNA containing biotinylated pseudouridine, which was then purified, quantified, and chemically fragmented at 95°C for 35 minutes. Fragmented biotinylated cRNA was hybridized in 200 μL hybridization buffer at 45°C for 16 hours to

[5] http://www.metasystems.de.
[6] http://www.Affymetrix.com.

Table 1. Characteristics of antibodies used for immunohistochemistry on TMAs

Protein	Antibody	Origin	Clone	Dilution
Aurora A (STK6/STK15/AURKA)	mmab	C. Prigent, Rennes	/	1/25
ERBB2	rpab	DakoCytomation	HercepTest	1/400
ER	mmab	Novocastra Laboratories	6F11	1/60
MIB1/Ki67	mmab	DakoCytomation	Ki67	1/100
Progesterone receptor	mmab	DakoCytomation	PgR 636	1/80

Abbreviations: mmab, mouse monoclonal antibody; rpab, rabbit polyclonal antibody.

microarrays contained over 47,000 transcripts and variants, including 38,500 well-characterized human genes. Automated washes of microarrays and staining with streptavidin-phycoerythrin were done according to the instructions of the manufacturer. Double signal amplification was done by biotinylated antistreptavidin antibody with goat-IgG blocking antibody. Scanning was done with Affymetrix GeneArray scanner and signals were quantified using Affymetrix GCOS software. All hybridization images were inspected for artifacts. Expression data was then analyzed by the RMA (Robust Multichip Average) method in R using Bioconductor and associated packages (19). RMA did the background adjustment, the quantile normalization, and finally the summarization of 11 oligonucleotides per gene.

Gene expression data analysis. Analysis of RNA expression data included classic unsupervised and supervised methods, and construction of a transcriptome correlation map. Before analysis, a filtering process removed from data set the genes with low and poorly measured expression as defined by expression value inferior to 100 units in all 55 breast cancer tissue samples, retaining 28,516 genes/expressed sequence tag (data are available in Supplementary Table S2).

Before unsupervised hierarchical clustering, a second filtering process removed the genes with low expression change across the 55 samples, as defined by SD inferior to 100 units (only for calculation of SD, values were floored to 100 because discrimination of expression variation in this low range cannot be done with confidence), retaining 12,157 probe sets (including 396 unique probe sets of chromosome 20). Hierarchical clustering of $\log_2$-transformed expression data was done with the Cluster program (20) using Pearson correlation as similarity metric and centroid linkage clustering. Results were displayed using the TreeView program (20).

Supervised analyses were applied to the 28,516 genes/expressed sequence tag to identify genes that discriminate between samples with and samples without amplification on 20q13, and between samples with and samples without *ZNF217* amplification. A discriminating score (DS) was calculated for each gene (21) as $DS = (M_1 - M_2) / (S_1 + S_2)$, where M_1 and S_1, respectively, represent mean and SD of expression levels of the gene in class I, and M_2 and S_2 in class II. Because of multiple hypotheses testing, confidence levels were estimated by 100 iterative random permutations of samples as previously described (2) and by computing the proportion of permutations in which the number of genes selected exceeds the observed number of genes.

We then applied an analysis similar to that used by Reyal et al. (22) to generate a transcriptome correlation map based on the correlated expression of neighboring genes in our 55 samples. Oligonucleotide probe sets were mapped to the genome according to the National Center for Biotechnology Information Ensemble database and the University of California San Francisco Genome Bioinformatics database Genome Browser.[7] From the 28,516 probe sets above selected after filtering, 678 were assigned to chromosome 20. Because some genes are represented by several probe sets on Affymetrix microarrays, which may introduce artifacts in analysis, we retained a unique probe set to each gene by selecting the probe set that presented the highest correlation with the median profile obtained with all corresponding probe sets. Such filtering retained 396 unique probe sets on chromosome 20, including 266 on the long arm, and 100 on the 20q13 region. To evaluate the correlation between the expression profile of each gene located on 20q and that of its neighbors, we computed a transcriptome correlation score for each probe set. This score was defined as the average of the correlation coefficients across all samples between RNA expression levels of the probe set and the RNA expression levels of each of the physically nearest 20 probe sets (10 centromeric and 10 telomeric). The significance threshold for the transcriptome correlation score was obtained by permutation tests in using the 1,000th quantile of the random distribution (i.e., the value for which 1 of 1,000 probe sets in the random data sets was above this value). A probe set with a score exceeding the threshold was considered as significantly correlated with its neighbors.

Biologically relevant subtypes of breast tumors have been defined using an "intrinsic set" of 552 genes (23). To assign a molecular subtype to our samples, we analyzed the 55 samples with the 400 genes common to this intrinsic set and our Affymetrix data for 28,516 filtered genes.

Reverse transcription. RNA extracted from frozen tissues was reverse transcribed in a final volume of 20 μL. Two micrograms of RNA were combined with 0.4 μL of random hexamers (Invitrogen, Carlsbad, CA) and 1 μL of deoxynucleotide triphosphates mix (10 mmol/L; Invitrogen). After incubating the mixture at 65°C for 5 minutes, the reaction was cooled at 4°C for at least 1 minute. After annealing, RNA was added to 2 μL of 10× reverse transcription buffer (Invitrogen), 4 μL of 25 mmol/L $MgCl_2$ (Invitrogen), 2 μL DTT (Invitrogen), and 1 μL RNase inhibitors (Invitrogen). The solution was incubated at 25°C for 2 minutes after which 1 μL of reverse transcriptase (Superscript II Invitrogen) was added. After incubating 10 minutes at 25°C and 50 minutes at 42°C, the reaction was stopped by heating to 70°C for 15 minutes.

Quantitative PCR. Quantitative reverse transcription-PCR (RT-PCR) analyses for *ZNF217*, *ARFGEF2*, *DDX27*, *DPM1*, and *TBP* (control) mRNA were done using the 7900 HT fast Real-Time PCR System instrument and software (Applied Biosystems, Foster City, CA). Primers and probes for the Taqman system were designed to meet specific criteria by using Primer Express software (Applied Biosystems) and were synthesized by Sigma-Proligo (Proligo LLC, Boulder, CO) for the primers and by Applied Biosystems for the probes. The 5′- and the 3′-end nucleotides of the probe were labeled with a reporter (FAM, 6-carboxy-fluorescein) and the quencher dye (TAMRA, 6-carboxy-tetramethylrhodamine). The sequences of the PCR primer pairs and fluorogenic probes used for each gene are shown in Supplementary Table S3. The precise amount of total RNA added to each reaction mix (based on absorbance) and its quality are both generally difficult to assess. Therefore, the relative expression level of the gene of interest was computed with respect to the internal standard *TBP* gene to normalize for variations in the quality of RNA and the amount of input cDNA.

[7] http://genome.ucsc.edu/.

C_t (threshold cycle) was used for quantification of the input target number and all experiments were done with duplicates for each data point. No C_t variation for the duplicate exceeded our threshold value of 1.5. For each experimental sample, the amount of target and endogenous reference was determined from a standard curve. The standard curve was constructed with 5-fold serial dilutions of cDNA (100-0.1 ng) from four breast carcinoma cell lines: JIMT1 for *ARFGEF2*, BT-474 for *DDX27*, UACC-812 for *DPM1*, and BT-483 for *ZNF217*. The level of expression of the target gene for each sample was calculated as $2^{-\Delta Ct\ sample}$, where $\Delta C_{t\ sample}$ was the difference in the number of cycles needed for the FAM fluorescence intensity for the target gene and the *TBP* reactions to reach a threshold value. All measurements were normalized using a calibrator sample, i.e., the HME1 human primary mammary epithelial cell line (Clontech, Mountain View, CA). Thus, the relative expression level of the gene of interest and *TBP* for each sample was $RE_{sample} = 2^{-(\Delta Ct\ sample\ -\ \Delta Ct\ HME1)}$, where $\Delta C_{t\ HME1}$ was the difference in the number of cycles needed for the FAM fluorescence intensities from the target gene and *TBP* reactions to reach a threshold value in the HME1 calibrator sample.

PCR was done with 1× Taqman Universal PCR Master Mix (Applied Biosystems manufactured by Roche, Branchburg, NJ), 300 nmol/L of primers, 200 nmol/L of the probe, and 2.5 μL of each appropriately diluted reverse transcription sample in 25 μL final reaction mixture. After 2 minutes incubation at 50°C to allow for uracyl *N*-glycosylate cleavage, AmpliTaq Gold was activated by an incubation for 10 minutes at 95°C. Each of the 40 PCR cycles consisted of 15 seconds of denaturation at 95°C and hybridization of probe and primers for 1 minute at 60°C.

Statistical analysis. The correlation of RNA expression data obtained by DNA microarrays and by quantitative RT-PCR was measured using Pearson's correlation and Spearman's rank correlation. Distributions of molecular markers and other categorical variables were compared using either the χ^2 or the Fisher's exact test. The follow-up was calculated from the date of diagnosis to the time of metastasis as first event or time of last follow-up for censored patients. The end point was the metastasis-free survival (MFS), calculated from the date of diagnosis, first distant metastasis being scored as an event. All other patients were censored at the time of the last follow-up, death, recurrence of local or regional disease, or development of a second primary cancer. Survival curves were derived from Kaplan-Meier estimates and compared by log-rank test. Statistical tests were two-sided at the 5% level of significance.

Results

Amplification of 20q13 genes in breast cancer. We determined the frequency of amplification of the 20q13 region by FISH on a breast cancer TMA. Using BAC clones, we prepared genomic probes for three selected loci (*MYBL2*, *STK6*, *ZNF217*; see Fig. 1A). Of the 685 tumors screened, 466 gave satisfactory FISH results. Examples are shown in Fig. 1B-C. We found 36 tumors (8%) with 20q13 amplification and 430 (92%) without amplification of any three loci. Of the 36 amplified cases, 53%, 86%, and 53% exhibited amplification of *MYBL2*, *ZNF217*, and *STK6*, respectively (Fig. 1A). *ZNF217* was the most frequently amplified locus, in agreement with previous studies (9). In 9 of the 36 samples (thereafter designated type 1), *ZNF217* was even amplified without amplification of the two other loci.

20q13 Amplification and histoclinical correlations. We compared gene amplification with histoclinical data. No correlation was found between global amplification of the 20q13 region and any studied histoclinical feature (data not shown). Two strikingly opposite data were observed depending on whether we considered tumors amplified for *ZNF217* locus only (type 1) or tumors with amplification at two or three loci (type 2). ZNF217-only amplified tumors were associated with the absence of lymph node metastasis ($P < 0.025$, Fisher's exact test), whereas type 2 tumors were associated with high histologic grade ($P < 0.005$, Fisher's exact test), the presence of lymph node metastasis ($P < 0.00001$, Fisher's exact test), and the expression of Aurora A protein (measured by immunohistochemistry; $P < 0.025$, Fisher's exact test; Table 2). Unfortunately, no reliable antibody was available for MYBL2 and ZNF217 proteins and we could not examine their expression.

With a median follow-up of 60 months after diagnosis (range, 2-151), analysis of MFS of patients with 20q13-amplified tumor confirmed these two opposite data. In the whole population of samples, ZNF217-only amplified tumors, found as associated with the absence of lymph node invasion, were paradoxically associated with a poor prognosis [5-year MFS 62% (35.53-100)]. Type 2 tumors, associated with lymph node invasion, were paradoxically associated with a good prognosis [5-year MFS 91% (80.2-100); $P = 0.053$, log-rank test; Fig. 2A]. Moreover, amplification at *ZNF217* was strongly associated with a poor outcome in the lymph node–negative population [5-year MFS 64% (33.84-100)] when compared with no amplification [5-year MFS 92.27% (87.87-95.6)], whereas amplification of at least two loci was strongly associated with a good prognosis in the lymph node–positive population [5-year MFS 89% (70.56-100), versus 55% (50.78-75.87) for nonamplified tumors; Fig. 2B-C].

Gene expression profiling of 20q13-amplified and nonamplified tumors. The different effects on prognosis suggested that the two types of amplified tumors might represent distinct entities. To better define these potential entities, we established their gene expression profiles on a pan-genome scale. Based on FISH and histoclinical correlations, we selected for analysis with whole-genome DNA microarrays 55 tumor samples with available RNA representing three different groups (nonamplified population: 36 cases; amplified population: 19 samples, including 7 type 1 cases and 12 type 2 cases).

Unsupervised hierarchical analysis. Before analysis, a filter procedure eliminated genes with uniform low expression or with low expression variation across the experiments, retaining 12,157 probe sets. Results of hierarchical clustering are shown in Fig. 3. The tumors displayed heterogeneous expression profiles, as reflected by the dendrogram branch length (Fig. 3B). Overall, they fell in two classes. The distribution of tumors among the two classes was correlated with the 20q13 amplification status. Class I (27 samples) included the majority of the 20q13-amplified tumors (14 of the 19 amplified samples, including six type 1 tumors and eight type 2 tumors), whereas class II contained only five 20q13-amplified tumors (one type 1 and four type 2) among 28 samples ($P = 0.011$, Fisher's exact test). A strong correlation existed between the two classes and the ER status of tumors, with more ER-positive tumors in class I compared with class II ($P < 0.0001$, Fisher's exact test).

Gene clustering revealed groups of coexpressed genes, some of which represented expression signatures corresponding to defined molecular subtypes (see colored bars on the right of Fig. 3A and zooms in Fig. 3C). Two major clusters were prominent in the classification of tumors. One of them (652 probe sets) included *ESR1*, which codes for ER-α, and several genes frequently described associated with ER-positive status (*GATA3*, *XBP1*, *SPDEF*, *KRT8*, *KRT18*, *CCND1*). This cluster

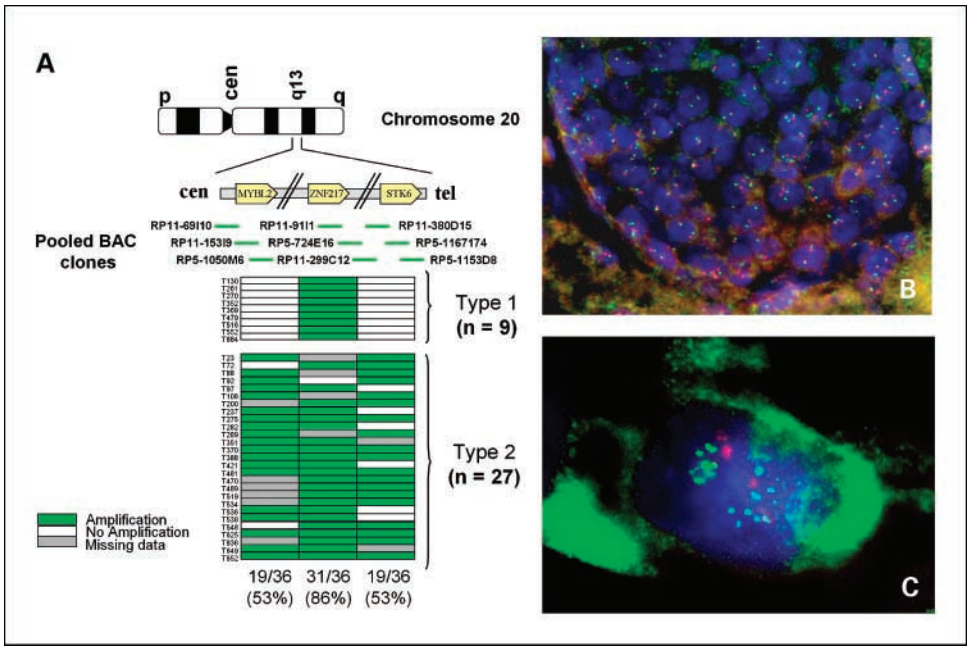

Fig. 1. Determination of 20q13 amplification using FISH analysis. *A*, schematic representation of tumor amplification status for each of the three loci selected as probes. Combinations of BAC pooled clones used for the FISH analysis are shown for each locus. Tumors were grouped together by similarity of amplification status: tumors without amplification, 430 cases; type 1, tumors with amplification of *ZNF217* only, nine cases; type 2, tumors with amplification of, at least, two genes, 27 cases. The amplification frequency is calculated for each locus among the 36 tumors exhibiting amplification. *B* and *C*, illustration of two-color FISH analyses of breast cancer nuclei using BAC pools as probes, representing centromere 20 (*red*) and *ZNF217* gene region (*green*). In (*B*), a tumor without amplification and in (*C*), a tumor with amplification of the *ZNF217* gene region.

was very similar to the previously reported luminal gene cluster (24) and was overexpressed in class I tumors overall compared with class II tumors. The other major cluster (483 probe sets) was overexpressed in class II compared with class I. This cluster was very similar to the previously reported basal gene cluster (*KRT5*, *KRT6A*, *KRT6B*, *KRT16*, *KRT17*, *ITGA6*, *TRIM29*, *S100A2*, *SLP1*, *ANXA8;* ref. 24). We also identified an ERBB2-related cluster prominent in tumors with overexpression of the ERBB2 protein measured by immunohistochemistry. This cluster included *ERBB2*, *GRB7*, *PERLD1*, *STARD3*, and *C17ORF37*, all of which map to chromosomal region 17q12 and belong to an *ERBB2* GES (2). The agreement between our data and previously described luminal, basal, and ERBB2-related clusters (24) suggested the validity of our results.

Interestingly, we identified a 20q13 gene cluster (43 probe sets) overexpressed in class I overall compared with class II. This gene cluster included *ZNF217* and several genes localized in 20q13 chromosomal region between *MYBL2* and *STK6/Aurora A* genes (*ADNP*, *ARFGEF2*, *YWHAB*, *UBE2v1*, *DDX27*, *STAU*, *CSE1L*, *DPM1*, and *ELMO2*).

Molecular subtypes of 20q13-amplified tumors. To assign specifically a molecular subtype to each tumor, we directly confronted our data with those of Sorlie et al. (24). Using an intrinsic gene set (552 genes), these authors have identified five molecular subtypes in tumors (luminal A, luminal B, basal, ERBB2-related, and normal-like). Of the 552 gene set, 400 overlapped with our 28,516 filtered probe sets. Hierarchical clustering of the available RNA expression data for these 400 genes in the 122 samples from Sorlie et al. (24) still discriminated the same five molecular subtypes (data not shown). This allowed us to define the typical expression profile of each subtype for this gene set (thereafter designated centroid). The centroid expression for each of the five subtypes was then computed as the median expression for each of the 400 genes in the corresponding samples. We then measured the correlation of each of our 55 samples with each centroid, and identified 24 luminal A, 2 luminal B, 6 ERBB2-related, 17 basal, and 2 normal-like samples (four samples were not assigned any subtype). Tumors are color-coded by subtypes in the dendrogram in Fig. 3B. As shown, the major subgrouping of tumors based on global hierarchical clustering was in agreement with the subtypes from which they were closer: Most luminal A tumors were in class I, whereas class II included all basal tumors. Interestingly, the 20q13-amplified tumors tended to be of luminal A subtype, with 13 of 19 cases ($P = 0.045$, Fisher's exact test).

Gene expression signatures. We next used supervised analyses based on the 28,516 probe sets representing 15,771 genes to identify the genes most specific of each type of 20q13 amplification.

To establish type 1 GES, we compared the seven samples with amplification of *ZNF217* (type 1) with 36 samples without 20q amplification. We identified a signature of 498 discriminator probe sets (see Supplementary Table S4), of which 102 were overexpressed and 396 were underexpressed in the amplified samples (probability that this number of genes would be selected by chance, 0). *ZNF217* was not identified in this signature, and no gene from chromosomal region between *MYBL2* and *STK6* was included ($P = 0.2024$, Fisher's exact test; Fig. 4A). Only four of the 498 probe sets were localized on the long arm of the chromosome 20 and they lied outside the *MYBL2-STK6* region (Table 3).

To establish type 2 GES, we compared the 12 tumor samples presenting at least two amplified loci (type 2) with the 36 nonamplified samples. We identified a signature of 379 probe sets representing 324 genes (see Supplementary Table S5) differentially expressed between the two groups (probability that this number of genes would be selected by chance, 0.01). Among the 379 probe sets, 149 were overexpressed and 230 were underexpressed in the amplified samples. In contrast to type 1 GES, this signature was enriched in genes localized on the long arm of the chromosome 20, with 25 probe sets (22 genes) of this arm (6.6%), whereas the 20q-specific probe sets represent only 1.5% of the 28,516 analyzed Affymetrix probe sets ($P < 0.0001$, Fisher's exact test; Fig. 4B). Among the

Table 2. Correlation between amplification at 20q13 loci and histoclinical data

Characteristics	No amplification	*ZNF217* amplification only (type 1)	Multiple genes amplification (type 2)	Statistical significance		
	No. patients (%)	No. patients (%)	No. patients (%)	*P**	*P*†	*P*‡
All cases	431 (92)	9 (2)	26 (6)			
Age (y)						
≤50	121 (28)	3 (33)	7 (27)	NS	NS	NS
>50	310 (72)	6 (67)	19 (73)			
Histologic type						
Ductal	286 (66)	7 (78)	23 (92)	NS	NS	NS
Lobular	62 (14)	0 (0)	1 (4)			
Other	84 (20)	2 (12)	1 (4)			
Pathologic tumor size						
pT_1	185 (43)	5 (55)	8 (31)	NS	NS	NS
pT_2	182 (43)	3 (33)	12 (46)			
pT_3	61 (14)	1 (12)	6 (23)			
SBR grade						
I	143 (33)	2 (22)	1 (4)	NS	<0.005	NS
II	192 (45)	4 (44)	13 (50)			
III	95 (22)	3 (34)	12 (46)			
Lymph node metastasis						
0	233 (54)	7 (78)	6 (23)	NS	<0.00001	0.025
1-3	130 (30)	1 (11)	7 (27)			
>3	66 (16)	1 (11)	13 (50)			
Peritumoral vascular invasion						
Absent	187 (44)	2 (22)	11 (42)	NS	NS	NS
Present	241 (56)	7 (78)	15 (58)			
ER						
Negative	81 (19)	3 (33)	4 (15)	NS	NS	NS
Positive	350 (81)	6 (67)	22 (85)			
Progesterone receptor						
Negative	131 (30)	5 (56)	6 (23)	NS	NS	NS
Positive	300 (70)	4 (44)	20 (77)			
ERBB2 receptor						
Negative	367 (87)	7 (78)	24 (89)	NS	NS	NS
Positive	54 (13)	2 (22)	3 (11)			
MIB1/Ki67						
>20	245 (80)	4 (57)	17 (77)	NS	NS	NS
≤20	62 (20)	3 (43)	5 (23)			
Aurora A/STK6						
Negative	155 (54)	2 (29)	6 (26)	NS	<0.025	NS
Positive	133 (46)	5 (71)	17 (74)			
Radiotherapy						
No	17 (4)	1 (12)	3 (11)	NS	NS	NS
Yes	424 (96)	8 (88)	24 (89)			
Hormone therapy						
No	201 (47)	6 (66)	9 (35)	NS	NS	NS
Yes	229 (53)	3 (33)	17 (65)			
Chemotherapy						
No	204 (47)	6 (66)	12 (48)	NS	NS	NS
Yes	227 (53)	3 (33)	14 (52)			
Metastasis						
No	369 (83)	6 (66)	25 (92)	NS	NS	NS
Yes	73 (17)	3 (33)	2 (8)			
Follow-up, mo, median (range)	60 (2-151)	57 (18-81)	61 (24-137)	NS	NS	NS

* *P* values for the comparison of histoclinical status between nonamplified tumors and tumors of type 1 were calculated using the Fisher's exact test.
† *P* values for the comparison of histoclinical status between nonamplified tumors and tumors of type 2 were calculated using the Fisher's exact test.
‡ *P* values for the comparison of histoclinical status between tumors of type 1 and tumors of type 2 were calculated using the Fisher's exact test.

22 genes were 12 genes from 20q13 located between *MYBL2* and *STK6* (*C20orf142*, *TDE1*, *YWHAB*, *NCOA3*, *ARFGEF2*, *DDX27*, *PARD6B*, *BCAS4*, *ADNP*, *DPM1*, *C20orf17*, *ZNF217*; Table 4 and Fig. 4B).

Transcriptome correlation map. Using a recently published method (22), we established a chromosomal "transcriptome correlation map" of chromosome 20 long arm to determine groups of neighboring genes that show correlated expression profiles. When analyzing the seven type 1 samples together with the 36 nonamplified samples (see Supplementary Fig. S1A), only 3.4% of the 20q genes were correlated with the expression of their neighbors, and no group of coexpression was identified in the 20q13 region. This result suggests the absence of driver gene of type 1 amplification. When analyzing the 12 type 2 samples and the 36 nonamplified samples (see Supplementary Fig. S1B), only 2.6% of the genes of the 20q region showed a significant correlation with their neighbors, and the 20q13 region contained no significantly correlated gene. This suggests that 20q13 amplification induces overexpression of isolated genes of the region.

Confirmation of gene expression measurements by quantitative RT-PCR. To confirm our expression results, we measured in a set of 24 tumors (seven type 1, nine type 2, and eight type 3) the expression of four discriminator genes overexpressed in type 2 tumors (*ZNF217*, *ARFGEF2*, *DDX27*, and *DPM1*) by quantitative RT-PCR. Expression values were correlated with microarray data. The correlation coefficients ranged from 0.68 to 0.87 for Pearson's correlation (ρ_p) and from 0.78 to 0.88 for Spearman's rank correlation (ρ_s; Fig. 5A). In type 1 tumors, expression of *ZNF217* was found in four cases but not in three cases. In contrast, the four genes were significantly expressed in type 2–amplified samples. Clustering of tumors according to the quantitative RT-PCR expression levels of the four genes correctly distinguished the type 2 but not the type 1 samples from the nonamplified samples (Fig. 5B). This showed that, in perfect agreement with DNA microarray data, expression of *ZNF217* was not associated with amplification in type 1 tumors, whereas the expression of the four genes was correlated with amplification in type 2 tumors.

Discussion

The 20q13 chromosomal region is amplified in 5% to 10% of breast cancers (9), as well as in other types of tumors (25, 26). It is also frequently amplified in breast cell lines (27). Several potential 20q13 oncogenes have been proposed, including *BCAS4* (28), *CAS/CSE1* (29), *CYP24* (30), *MYBL2* (27), *NABC1* (31, 32), *NCOA3/AIB1* (26), *STK6/AURKA/Aurora A* (33), *TDE1* (34), and *ZNF217* (35, 36). Aurora A is a very good candidate because its overexpression induces centrosome amplification and aneuploidy (10). ZNF217, a Kruppel-like zinc finger transcription factor, also has oncogenic potential; it can immortalize breast epithelial cells and can suppress cell

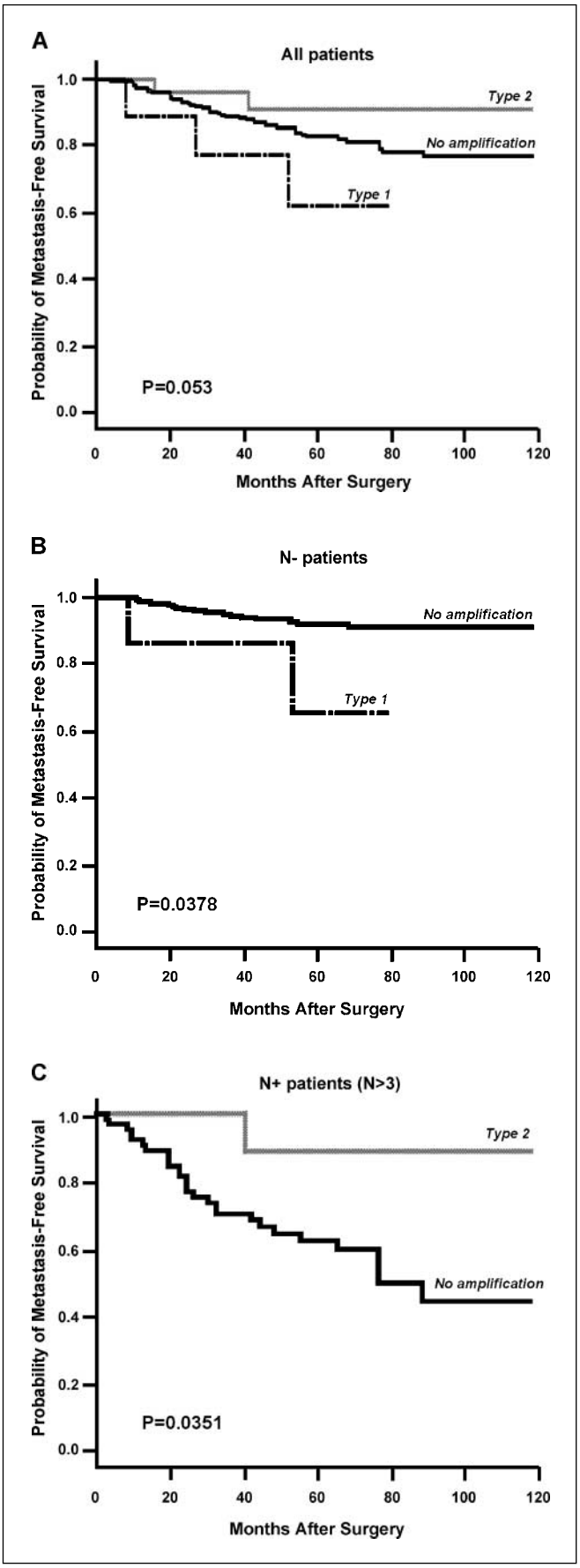

Fig. 2. 20q amplification and survival in breast cancer. Kaplan-Meier analyses of the MFS according to 20q13 amplification. *A*, survival of all 466 patients according to 20q amplification status (tumors without amplification, or amplification of type 1 and 2). *B*, survival of patients with node-negative cancer according to amplification of type 1 versus tumors without amplification. *C*, survival of patients with more than three involved axillary lymph nodes according to amplification of type 2 versus tumors without amplification. All *P* values are calculated by the log-rank test.

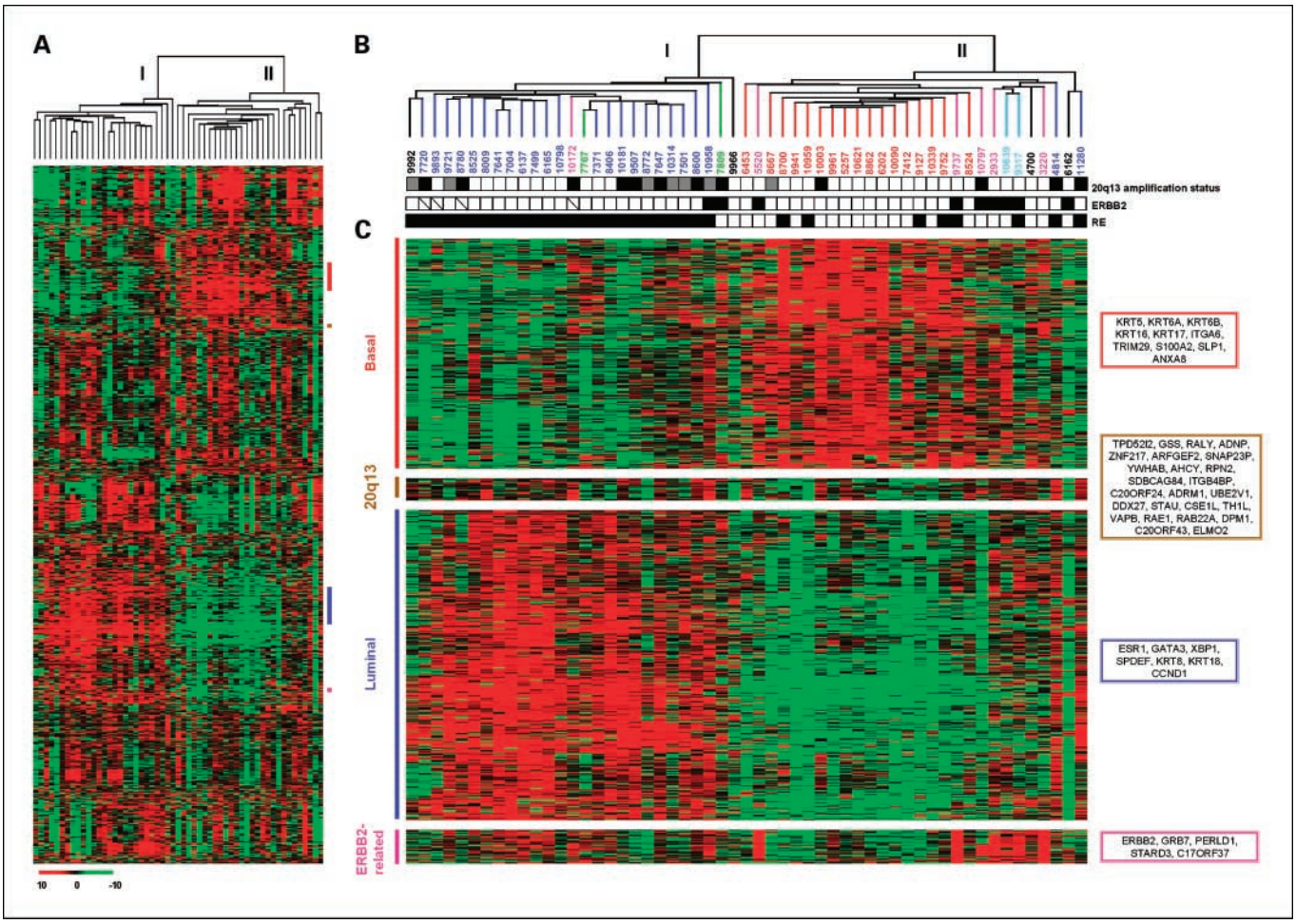

Fig. 3. Unsupervised hierarchical clustering of expression data. Total RNAs from 55 breast cancer tissue samples were hybridized to Affymetrix U133 Plus 2.0 human oligonucleotide microarrays. After filtering, analysis was focused on 12,517 probe sets. *A*, hierarchical clustering of 55 breast cancer samples based on mRNA expression levels of 12,517 probe sets. Each row represents a gene, each column a sample. The $\log_2$-transformed expression level of each gene in a single sample is relative to its median abundance across all samples and is depicted according to the color scale (*bottom*). The dendrogram of samples (*above matrix*) represents overall similarities in gene expression profiles. Two major classes of samples (I and II) are defined. Colored bars on the right, locations of four gene clusters of interest. *B*, dendrogram and molecular characteristics of samples. Branches of the dendrogram are color-coded according to the strongest correlation of our 55 tissue samples to five subtype centroïds described by Sorlie et al. (24): dark blue, luminal A subtype; light blue, luminal B; red, basal; pink, ERBB2-related; green, normal-like; black, samples with low correlation (inferior to 0.15) to any centroid. Under the dendrogram, some relevant features of samples are represented according to a color ladder (unavailable: oblique feature): 20q amplification status (tumors without 20q amplification: *white*; type 1 tumors with *ZNF217* amplification: *gray*; type 2 tumors with amplification of at least two loci: *black*); immunohistochemistry status for ER and ERBB2 (positive: *black*; negative: *white*). *C*, expanded view represents four selected gene clusters named from top to bottom on the left: Basal (*red bar*); 20q13 (*brown bar*); luminal (*dark blue bar*); and ERBB2-related (*pink bar*). Some genes included in these clusters are referenced by their HUGO abbreviation as used in Entrez Gene.

death by acting on the AKT pathway (35). The structure (6, 7) and boundaries (37, 38) of the 20q13 amplification have been defined but many uncertainties remain. The amplification seems complex, and multiple subregions and several genes seem to participate to the event (9, 39). Intriguingly, no extensive analysis of gene expression has yet been conducted on the 20q13 amplification or on the proposed oncogenes.

Here, we have established the gene expression profiles of breast tumors with 20q13 amplification. We found 36 amplified tumors out of a series of 466 by FISH using probes for three 20q13 loci, *MYBL2*, *ZNF217*, and *STK6/Aurora A*. This rate of amplification (8%) was in agreement with literature (9). We observed two types of amplification. In the first type, the *ZNF217* locus was amplified but we did not detect amplification at the two other loci. In the second type, the amplified region seemed larger and involved two or three loci. The two types showed different gene expression profiles and were associated with different histoclinical features including lymph node status and survival.

Tanner et al. (8) also distinguished two types of 20q13 alteration, one with a low level of amplification, and a second with a high level of amplification and described three subregions of amplification (6). Their study showed that the 20q13 amplification was associated with aneuploidy and high grade in breast cancer (8). High level of 20q13 amplification was associated with short disease-free survival in lymph node–negative patients. The same is true for other types of cancer (26, 40). The correlation of *STK6* amplification and high grade has been confirmed in other studies (41) but the effect on survival was not observed. This discrepancy might be explained by the fact that the 20q13-amplified tumors are heterogeneous. We found that tumors with amplification at

ZNF217 only were more frequently node-negative tumors, but were paradoxically associated with a poor prognosis, whereas tumors with amplification at two or three loci were more frequently node-positive tumors paradoxically associated with a good prognosis. Because these observations were obtained in relatively small number of amplified samples, which represent 8% of breast cancers, they need to be validated in larger series.

Our profiling of 20q13-amplified tumors confirmed the hypothesis of two types of amplification. Type 1 (ZNF217 only) and type 2 had distinct GESs. However, due to the low number of cases, we were not able to perform direct comparison of the two types. Type 1 tumors showed no significant overexpression of any gene from the chromosomal region located between *MYBL2* and *STK6*, including *ZNF217* itself. In some cases, expression of *ZNF217* was detected by DNA microarrays and quantitative RT-PCR but without any statistical correlation with amplification. This suggests that 20q13 amplification in type 1 tumors may result from a phenomenon not linked to a particular 20q gene. This putative phenomenon is apparently associated with a poor prognosis. Alternatively, given the evidence for the oncogenic potential of ZNF217, which promotes immortalization of primary cells (35) and induces cell survival (42), a role for this gene cannot be ruled out. One hypothesis may be that some tumors overexpress *ZNF217* as an early antiapoptotic, anti-stress mechanism that may later disappear when other mechanisms take over. Type 2 tumors showed overexpression of a number of genes dispersed throughout the region, including several already proposed oncogenes such as *TDE1*, *NCOA3*, *BCAS4*, and *ZNF217*. This suggests that amplification may be driven by several genes rather than by a particular gene. Quantitative RT-PCR confirmed the good correlation between amplification and expression of *ZNF217* in our type 2 samples. Collins et al. (31) reported similar results in a panel of 11 tumor samples, in which a majority of type 2 samples may have been included. A possibility is that type 1 amplification represents an early stage of 20q amplification

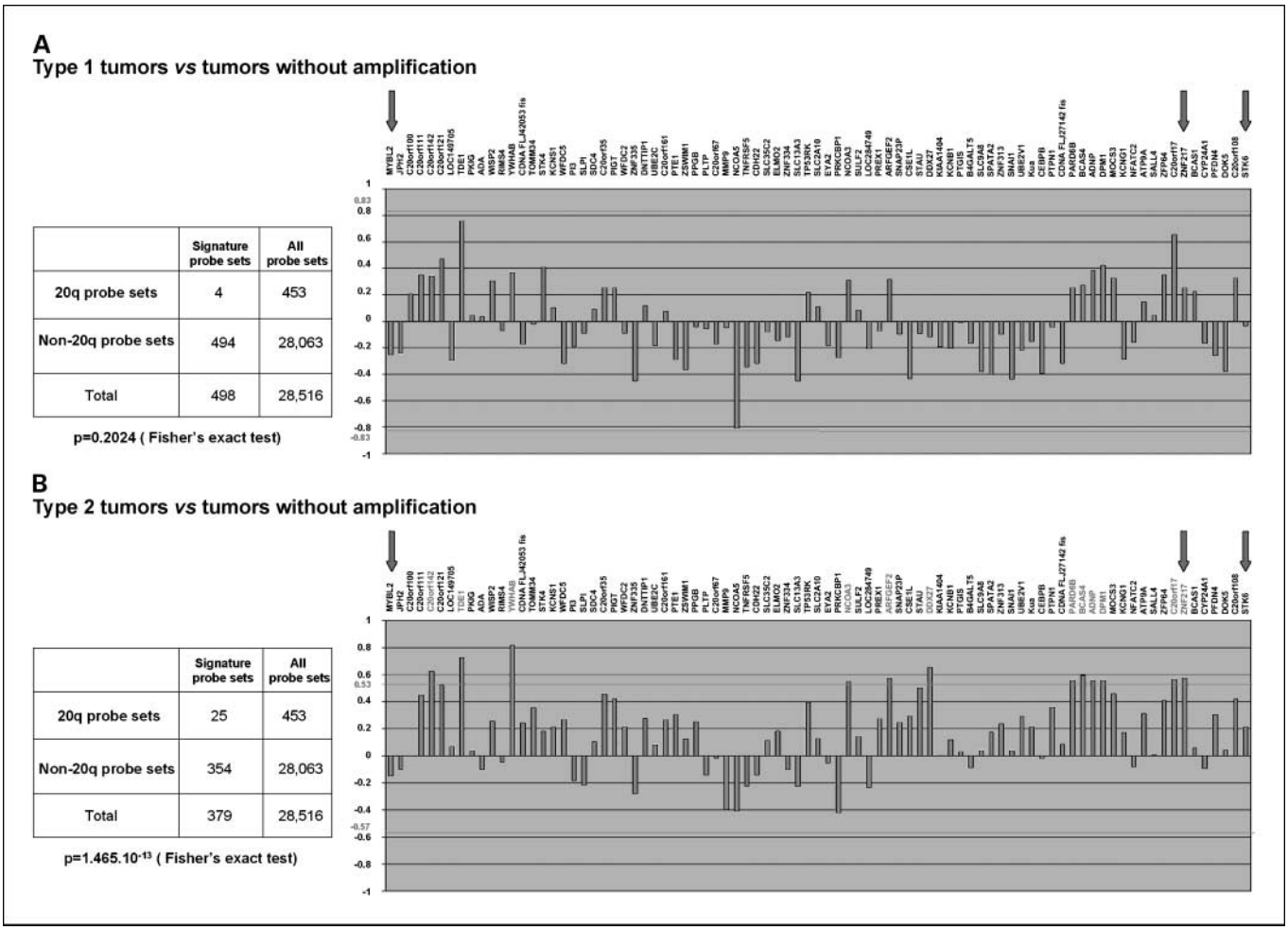

Fig. 4. Gene expression signatures. *A*, results of supervised analysis between type 1 tumors and tumors without amplification. Left, in comparison with whole genome, the proportion of probe sets of the 20q region in the signature was not different ($P = 0.2024$, Fisher's exact test). Right, DS of genes located in the region between *MYBL2* and *STK6* (*bars*). On top of the histogram, genes are referenced by their HUGO abbreviation as used in Entrez Gene. Two lines, significant threshold (equal to 0.53 and −0.57) for discriminator genes. No discriminator gene could be defined in this region. Vertical gray-colored arrows, location of the 20q13 genes analyzed by FISH: *MYBL2, ZNF217*, and *STK6*. *B*, results of supervised analysis between type 2 tumors and tumors without amplification. Left, this signature was enriched in genes localized on 20q13 region with 25 probe sets (22 genes) of this arm (6.6%), whereas the 20q-specific probe sets represent only 1.5% of the 28,516 Affymetrix probe sets ($P < 0.0001$, Fisher's exact test). Right, legend similar to (*A*); significant threshold equal to 0.83 and −0.83. Twelve genes located between *MYBL2* and *STK6* had a significant differential expression between type 2 samples and samples without amplification.

Table 3. Four 20q genes whose expression discriminates ZNF217-amplified (type 1) tumors from nonamplified tumors

Symbol	Description	RefSeq transcript ID*	Probe set ID†	Cytoband	Chromosome start (bp)	DS‡
STX16§	Syntaxin 16	NM_001001433	221499_s_at	20q13.32	56,659,743	1.48
EMILIN3§	Elastin microfibril interfacer 3	NM_052846	228307_at	20q11.2-q12	39,422,020	1.05
C20orf40	Chromosome 20 open reading frame 40	—	231200_at	20q13.33	60,141,738	−1.11
ITCH	Itchy homologue E3 ubiquitin protein ligase (mouse)	NM_031483	236235_at	20q11.22-q11.23	32,561,936	1.50

NOTE: Information was collected on the Affymetrix NetAffx web site.
*RefSeq Transcript ID: mRNA accession number.
† Probe set ID: identification Affymetrix of the probe set.
‡ DS: results of the supervised analysis.
§Genes common to type 1 and type 2 GES.

and that cases with amplification and expression of *ZNF217* may evolve in type 2 amplification with the participation of neighbor genes. Further analyses using tiling-path array comparative genomic hybridization on many cases should determine whether distinct 20q13 amplicons can be recurrently identified in type 2 tumors. The region may also be "packaged" in a single amplicon as to contain only amplified and overexpressed selected genes (43, 44). The participation of

Table 4. Twenty-two 20q genes whose expression discriminates type 2 amplified tumors from nonamplified tumors

Symbol	Description	RefSeq transcript ID*	Probe set ID†	Cytoband	Chromosome start (bp)	DS‡
C20orf44	Chromosome 20 open reading frame 44	NM_018244	217935_s_at	20q11.23	33,353,784	0.56
C20orf52	Chromosome 20 open reading frame 52	NM_080748	224972_at	20q11.23	33,750,679	0.54
RNPC2	RNA-binding region (RNP1, RRM) containing 2	NM_004902	227223_at	20q11.23	33,790,655	0.63
CTNNBL1	Catenin, β like 1	NM_030877	221021_s_at	20q11.23-q12	35,841,098	0.72
EMILIN3§	Elastin microfibril interfacer 3	NM_052846	228307_at	20q11.2-q12	39,422,020	0.69
C20orf142	Chromosome 20 open reading frame 142	XM_371399	226805_at	20q13.12	42,364,894	0.62
TDE1	Tumor differentially expressed 1	NM_006811	211769_x_at	20q13.1-13.3	42,562,241	0.73
YWHAB	14-3-3 Protein β/α-tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein, β polypeptide	NM_003404	208743_s_at	20q13.1	42,947,763	0.82
NCOA3	Nuclear receptor coactivator 3	NM_006534	209060_x_at	20q12	45,564,027	0.55
ARFGEF2	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)	NM_006420	222518_at	20q13.13	46,971,833	0.57
DDX27	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	NM_017895	219108_x_at	20q13.13	47,283,289	0.66
PARD6B	Par-6 partitioning defective 6 homologue β (*C. elegans*)	NM_032521	235165_at	20q13.13	48,801,685	0.56
BCAS4	Breast carcinoma amplified sequence 4	NM_001010974	228787_s_at	20q13.13	48,926,606	0.60
ADNP	Activity-dependent neuroprotector	NM_015339	226426_at	20q13.13	48,938,991	0.56
DPM1	Dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit	NM_003859	202673_at	20q13.13	48,984,811	0.56
C20orf17	Chromosome 20 open reading frame 17	NM_173485	220213_at	20q13.2	51,538,901	0.56
ZNF217	Zinc finger protein 217	NM_006526	203739_at	20q13.2	51,617,019	0.57
STX16§	Syntaxin 16	NM_001001433	221499_s_at	20q13.32	56,659,743	0.67
ATP5E	ATP synthase, H^+ transporting, mitochondrial F1 complex, epsilon subunit	NM_001001977	217801_at	20q13.3	57,037,137	0.54
PPP1R3D	Protein phosphatase 1, regulatory subunit 3D	NM_006242	204555_s_at	20q13.3	57,947,029	0.69
PSMA7	Proteasome (prosome, macropain) subunit, α type, 7	NM_002792	201114_x_at	20q13.33	60,145,185	0.56
GTPBP5	GTP binding protein 5 (putative)	NM_015666	1568822_at	20q13.33	60,200,784	0.62

NOTE: Information was collected on the Affymetrix NetAffx web site.
*RefSeq Transcript ID: mRNA accession number.
† Probe set ID: identification Affymetrix of the probe set.
‡ DS: results of the supervised analysis.
§Genes common to type 1 and type 2 GES.

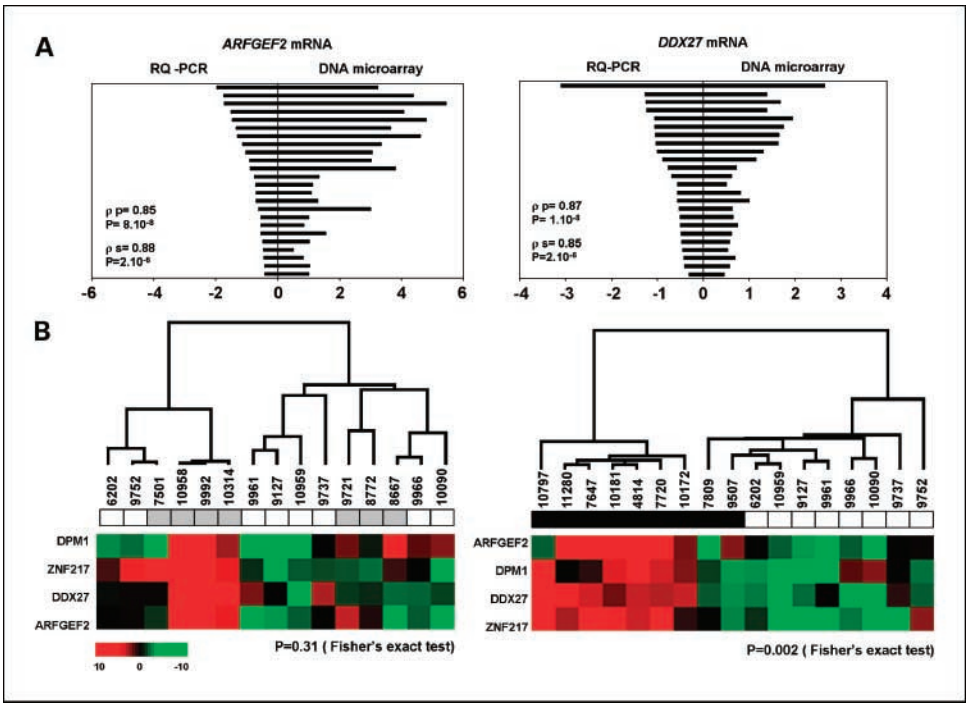

Fig. 5. Gene expression analysis by quantitative RT-PCR. *A*, comparison of mRNA expression levels of *ARFGERF2* and *DDX27* measured by DNA microarrays (*right*) and quantitative RT-PCR (*left*) in 24 tumors. Results for each tumor (from top to bottom) are represented as opposite bars. Correlations are measured using Pearson's correlation (ρ_p) and Spearman's rank correlation (ρ_s). For *ARFGERF2*: $\rho_p = 0.85$ ($P = 8 \times 10^{-8}$) and $\rho_s = 0.88$ ($P = 2 \times 10^{-6}$); for *DDX27*: $\rho_p = 0.87$ ($P = 1 \times 10^{-8}$) and $\rho_s = 0.85$ ($P = 2 \times 10^{-6}$). *B*, hierarchical clustering of 24 tumors and four genes (*ZNF217, ARFGEF2, DDX27*, and *DPM1*) based on expression data measured by quantitative RT-PCR. The 20q13 amplification status of tumors is indicated under the dendrogram. Legend is similar to Fig. 3A. Note that the clustering does not sort type 1 and nonamplified tumors ($P = 0.31$; *left*), whereas it correctly sorts type 2 and nonamplified tumors ($P = 0.002$; *right*).

Aurora A in type 2 amplification is highly suspected. However, *Aurora A* is also overexpressed without amplification in a high proportion of breast carcinomas (45). This explains why the gene is not identified as discriminator by our supervised analysis. In the population of patients with lymph node invasion, who are generally treated with chemotherapy, these tumors might be associated with a better prognosis due to a high level genome instability and response to treatment. The different prognosis of type 1 and type 2 tumors did not depend on their molecular subtypes because both tended to be luminal A tumors. This is in agreement with existing models (46). This subtype is generally of better prognosis than the other subtypes (24). The 20q13 amplification status might be used to further refine subclassification and prognosis of luminal A cancers.

Acknowledgments

We thank F. Birg, D. Maraninchi, and C. Mawas for encouragements, and C. Chabannon for the management of the biobank samples.

References

1. Al-Kuraya K, Schraml P, Torhorst J, et al. Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Res 2004;64:8534–40.
2. Bertucci F, Borie N, Ginestier C, et al. Identification and validation of an ERBB2 gene expression signature in breast cancers. Oncogene 2004;23:2564–75.
3. Garcia MJ, Pole JC, Chin SF, et al. A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. Oncogene 2005;24:5235–45.
4. Prentice LM, Shadeo A, Lestou VS, et al. NRG1 gene rearrangements in clinical breast cancer: identification of an adjacent novel amplicon associated with poor prognosis. Oncogene 2005;24:7281–9.
5. Gelsi-Boyer V, Orsetti B, Cervera N, et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res 2005;3:655–67.
6. Tanner MM, Tirkkonen M, Kallioniemi A, et al. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. Cancer Res 1996;56:3441–5.
7. Collins C, Volik S, Kowbel D, et al. Comprehensive genome sequence analysis of a breast cancer amplicon. Genome Res 2001;11:1034–42.
8. Tanner MM, Tirkkonen M, Kallioniemi A, et al. Amplification of chromosomal region 20q13 in invasive breast cancer: prognostic implications. Clin Cancer Res 1995;1:1455–61.
9. Hodgson JG, Chin K, Collins C, Gray JW. Genome amplification of chromosome 20 in breast cancer. Breast Cancer Res Treat 2003;78:337–45.
10. Zhou H, Kuang J, Zhong L, et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. Nat Genet 1998; 20:189–93.
11. Giet R, Petretti C, Prigent C. Aurora kinases, aneuploidy and cancer, a coincidence or a real link? Trends Cell Biol 2005;15:241–50.
12. Ginestier C, Charaffe-Jauffret E, Bertucci F, et al. Distinct and complementary information provided by use of tissue and cDNA microarrays in the study of breast tumor markers. Am J Pathol 2002;161:1223–33.
13. Chin SF, Daigo Y, Huang HE, et al. A simple and reliable pretreatment protocol facilitates fluorescent *in situ* hybridisation on tissue microarrays of paraffin wax embedded tumour samples. Mol Pathol 2003; 56:275–9.
14. Huang HE, Chin SF, Ginestier C, et al. A recurrent chromosome breakpoint in breast cancer at the NRG1/neuregulin 1/heregulin gene. Cancer Res 2004;64:6840–4.
15. van der Burg M, Poulsen TS, Hunger SP, et al. Split-signal FISH for detection of chromosome aberrations in acute lymphoblastic leukemia. Leukemia 2004;18: 895–908.
16. Cremet JY, Descamps S, Verite F, Martin A, Prigent C. Preparation and characterization of a human aurora-A kinase monoclonal antibody. Mol Cell Biochem 2003;243:123–31.
17. Jacquemier J, Ginestier C, Rougemont J, et al. Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. Cancer Res 2005;65: 767–79.
18. Theillet C, Adelaide J, Louason G, et al. FGFRI and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. Genes Chromosomes Cancer 1993;7:219–26.
19. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4: 249–64.
20. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95: 14863–8.
21. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531–7.
22. Reyal F, Stransky N, Bernard-Pierrot I, et al. Visualizing chromosomes as transcriptome correlation maps: evidence of chromosomal domains containing co-expressed genes—a study of 130 invasive ductal breast carcinomas. Cancer Res 2005;65:1376–83.

23. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
24. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23.
25. Bischoff JR, Anderson L, Zhu Y, et al. A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. EMBO J 1998;17:3052–65.
26. Tanner MM, Grenman S, Koul A, et al. Frequent amplification of chromosomal region 20q12-13 in ovarian cancer. Clin Cancer Res 2000;6:1833–9.
27. Forozan F, Mahlamaki EH, Monni O, et al. Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. Cancer Res 2000;60:4519–25.
28. Barlund M, Monni O, Weaver JD, et al. Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo amplification, overexpression, and fusion in breast cancer. Genes Chromosomes Cancer 2002;35:311–7.
29. Brinkmann U, Gallo M, Polymeropoulos MH, Pastan I. The human CAS (cellular apoptosis susceptibility) gene mapping on chromosome 20q13 is amplified in BT474 breast cancer cells and part of aberrant chromosomes in breast and colon cancer cell lines. Genome Res 1996;6:187–94.
30. Albertson DG, Ylstra B, Segraves R, et al. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet 2000;25:144–6.
31. Collins C, Rommens JM, Kowbel D, et al. Positional cloning of ZNF217 and NABC1: genes amplified at 20q13.2 and overexpressed in breast carcinoma. Proc Natl Acad Sci U S A 1998;95:8703–8.
32. Beardsley DI, Kowbel D, Lataxes TA, et al. Characterization of the novel amplified in breast cancer-1 (NABC1) gene product. Exp Cell Res 2003;290:402–13.
33. Sen S, Zhou H, White RA. A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. Oncogene 1997;14:2195–200.
34. Player A, Gillespie J, Fujii T, et al. Identification of TDE2 gene and its expression in non-small cell lung cancer. Int J Cancer 2003;107:238–43.
35. Nonet GH, Stampfer MR, Chin K, Gray JW, Collins CC, Yaswen P. The ZNF217 gene amplified in breast cancers promotes immortalization of human mammary epithelial cells. Cancer Res 2001;61:1250–4.
36. Rooney PH, Boonsong A, McFadyen MC, et al. The candidate oncogene ZNF217 is frequently amplified in colon cancer. J Pathol 2004;204:282–8.
37. Tanner MM, Tirkkonen M, Kallioniemi A, et al. Increased copy number at 20q13 in breast cancer: defining the critical region and exclusion of candidate genes. Cancer Res 1994;54:4257–60.
38. Weiss MM, Snijders AM, Kuipers EJ, et al. Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray comparative genomic hybridization. J Pathol 2003;200:320–6.
39. Bar-Shira A, Pinthus JH, Rozovsky U, et al. Multiple genes in human 20q13 chromosomal region are involved in an advanced prostate cancer xenograft. Cancer Res 2002;62:6803–7.
40. Aust DE, Muders M, Kohler A, et al. Prognostic relevance of 20q13 gains in sporadic colorectal cancers: a FISH analysis. Scand J Gastroenterol 2004;39:766–72.
41. Royce ME, Xia W, Sahin AA, et al. STK15/Aurora-A expression in primary breast tumors is correlated with nuclear grade but not with prognosis. Cancer 2004;100:12–9.
42. Huang G, Krig S, Kowbel D, et al. ZNF217 suppresses cell death associated with chemotherapy and telomere dysfunction. Hum Mol Genet 2005;14:3219–25.
43. Volik S, Zhao S, Chin K, et al. End-sequence profiling: sequence-based analysis of aberrant genomes. Proc Natl Acad Sci U S A 2003;100:7696–701.
44. Volik S, Raphael BJ, Huang G, et al. Decoding the fine-scale structure of a breast cancer genome and transcriptome. Genome Res 2006;3:3.
45. Tanaka T, Kimura M, Matsunaga K, Fukada D, Mori H, Okano Y. Centrosomal kinase AIK1 is overexpressed in invasive ductal carcinoma of the breast. Cancer Res 1999;59:2041–4.
46. Simpson PT, Reis-Filho JS, Gale T, Lakhani SR. Molecular evolution of breast cancer. J Pathol 2005;205:248–54.

Clinical Cancer Research

Prognosis and Gene Expression Profiling of 20q13-Amplified Breast Cancers

Christophe Ginestier, Nathalie Cervera, Pascal Finetti, et al.

Clin Cancer Res 2006;12:4533-4544.